Missing Non-Voters and Misweighted Samples: Explaining the 2015 Great British Polling Miss

Jonathan Mellon

Nuffield College, University of Oxford

British Election Study

Christopher Prosser

University of Manchester

British Election Study

The pre-election polls for the 2015 UK General Election missed the final result by a considerable margin: underestimating the Conservative Party and overestimating Labour. We analyse evidence for five theories of why the polls missed using British Election Study data. We find limited evidence for systematic vote intention misreporting, late swing, systematically different preferences among "don't knows" or differential turnout of parties' supporters. By comparing the BES face-to-face probability sample and online panel, we show that the online survey's polling error is primarily caused by under-sampling non-voters, then weighting respondents to represent the general population. Consequently, demographic groups with a low probability of voting are over-weighted within the voter subsample. Finally, we show that this mechanism is likely partially responsible for the over-estimate of the Liberal Democrats in 2010, showing that this is a longstanding problem with British polls.

Note: This is a draft paper. The results and conclusions may change before publication, as new data becomes available and additional analysis is conducted. Comments are welcome. Contact us at jonathan.mellon@nuffield.ox.ac.uk and chris.prosser@manchester.ac.uk

The last few years have not been kind to political pollsters across the world. Opinion polls were well wide of the mark in the U.S. midterms, the Israeli general election, and the Greek, Scottish and Irish referendums. One of the most notable polling misses was the 2015 UK General Election – the worst disaster for British pollsters in nearly a quarter of a century. Pre-election polling put Labour and the Conservatives in a dead heat (British Polling Council 2015), with neither party likely to win an outright majority. Labour was seen by many as having the best chances of forming the next government (Fisher and Lewis-Beck 2016). The actual results came as a surprise: the Conservatives led Labour by 6.3 percentage points and won enough seats to form an overall majority. The polls' inaccuracy had important consequences for the election campaign, with media focusing on potential coalition negotiations and the Conservative campaign focusing heavily on the threat of an SNP/Labour coalition (Green, Jennings, and Mellon 2015; Cowley and Kavanagh 2016). Expectations of a hung parliament and potential coalition configurations also seem to have influenced people's vote choices (Green, Fieldhouse, and Prosser 2015).

Several competing explanations have been put forward for the large polling errors on May 7th. This article examines five plausible explanations of the polling errors, looking at the evidence for each of these possible explanations from the British Election Study (BES).

1 Possible causes of polling errors

The first possible explanation is late swing. The "late swing" hypothesis states that the Conservative vote was underestimated because voters changed their minds at the end of the campaign, after pollsters finished contacting them. Investigations suggest that late swing made a modest contribution to the polling errors in 1992 election (Market Research Society Working Party 1994; Crewe 1992; Jowell et al. 1993a) and a larger contribution to the errors in 1970 (Abrams 1970).

The second possible explanation is differential turnout. This explanation states that some parties' supporters are less likely to go to the polls on Election Day than others. A large body of evidence shows that voters systematically overstate their likelihood of turning out to vote and over-report having voted after the fact (Burden 2000; Karp and Brockington 2005; Holbrook and Krosnick 2010; Presser and Traugott 1992; Górecki 2011). Consequently, pollsters must predict which respondents will actually go to the polls. At the 2015 British election, pollsters generally expected increased turnout, suggesting that their filters included too many respondents in their samples who ultimately did not vote. For instance Ipsos MORI reported that they expected a turnout of 72-74%, compared to the actual figure of 66.1% (Ipsos MORI 2015). Miscalibrated turnout models have been identified as one of the important reasons behind the under-estimate of the Republican vote share in the 2010 U.S. midterm elections (Pew Research Center 2016). If Conservative supporters were more likely to turn out to vote at the 2015 election than Labour supporters, this could explain the discrepancy between the pre-election polls and the actual result.

The third possible explanation of the polling miss is that respondents who said that they "don't know" who they would vote for chose differently on Election Day than respondents who declared a vote intention. This could be either because the "don't knows" had not yet made up their minds or because they were unwilling to reveal their vote choice. Some evidence suggests that differential voting among 'don't knows' contributed to the 1992 polling miss, with about 9% more of those saying don't know voting Conservative (27%) than Labour (18%) (Jowell et al. 1993b). A related idea is the 'spiral of silence' effect, whereby voters who support a party they believe is unpopular are reluctant to

share their views. This reluctance could manifest either by respondents refusing to reveal their vote intention or by not responding to surveys (Noelle-Neumann 1993). In the British context the spiral of silence is said to result in Conservative supporters differentially responding to opinion polls, distorting the polls towards Labour and is often suggested as one of the primary causes of the 1992 polling disaster (Crewe 1992; Turner and Sparrow 1997; Jowell et al. 1993b; Curtice, Sparrow, and Turner 1997). More recently the spiral of silence was cited as one of the possible reasons for the over-estimation of the Liberal Democrats in 2010 (Boon and Curtice 2010).

The fourth potential explanation is that respondents said one thing to pollsters while intending to do something else (Singh 2015). We call this problem "vote intention misreporting". There is some evidence of this type of misreporting in other elections (Streb et al. 2008) and vote intention misreporting is commonly cited as the reason for the 1992 polling miss,[1] although the extent to which voters were openly lying to the pollsters has been disputed (Crewe 1992; Worcester 1996). One possible factor that might contribute to vote intention (or choice) misreporting is priming through question order. Some pollsters suggested that question ordering is a possible cause of vote intention misreporting , including Labour's private pollster, James Morris and YouGov president Peter Kellner (Pedley 2015; Kellner 2015). In standard YouGov polls, substantial numbers of "inconsistent" respondents say they will vote Labour but rate the Conservatives higher on the economy. However there are fewer "inconsistent" voters in the BES campaign survey (also conducted by YouGov). Kellner suggests this difference may be because the BES asks for a respondent's vote intention after asking about party leaders and issues. This theory hypothesises that some Labour intenders who were "inconsistent" during the campaign switched to the Conservatives on Election Day. However, this explanation assumes that voters are consistent in their evaluations and vote choice when they actually vote.

The fifth potential explanation says that the polls may not have achieved representative samples of the electorate. Researchers have long recognised that representative samples are a requirement for reliable measurements of public opinion (Wilks 1940). However the reality of low response rates for most polls means that pollsters start off with highly unrepresentative samples, and have to apply substantial weighting in order to achieve representativeness. Weighting has become increasingly important as response rates have declined (Curtin, Presser, and Singer 2005; Keeter et al. 2006) and the use of non-randomly selected Internet panels has increased (Couper and Miller 2008; Chang and Krosnick 2009). However, given that the problem of achieving representative samples is longstanding and there were few large changes to most pollsters' methodology since the previous election, if unrepresentative samples are to blame for the polling miss, this raises the question of why the polls failed so badly in 2015 and not at previous elections.[2]

[1] After the 1992 polling miss the writer and Labour party donor Robert Harris wrote: 'I have reached the reluctant conclusion that ours is a nation of liars. People lied about their intentions up to the moment of voting, and went on lying even as they left the polling stations . . . The cynics were right after all. People may say they would prefer better public services, but in the end they will vote for tax cuts. At least some of them had the decency to feel too ashamed to admit it.' Quote in Crewe (1992, 489)

[2] This is not to suggest that the polls have been perfect at previous elections. British polls display a regular tendency for overestimating the Labour share of the vote and the 2010 election was particularly notable for the large estimate of the Liberal Democrat share of the vote. However in terms of the absolute error of the polls and the political consequences of misestimating the Conservative-Labour lead, the 2015 polling miss is the worst performance for British pollsters since at least the 1992 election.

After examining each of these theories we conclude that only one played an important role in explaining the Great British polling miss: representative samples. Our analysis shows that polls oversample politically engaged voters who turn out to vote at much higher rates than the average British person. Although pollsters *de facto* sample from the electorate (due to the small number of non-voters they actually reach), British polls weight their samples to population-level targets. This process means that the under-sampling of non-voters and the disengaged leads to the overweighting of respondents who share some demographic characteristics with non-voters. In 2015, these overweighted respondents were disproportionally Labour supporting, inflating the estimate of the Labour vote. Although the problem of representativeness is longstanding, our analysis suggests that in the 2010 election these respondents were disproportionally Liberal Democrats. The collapse of support for the Liberal Democrats after entering into coalition with the Conservative and the flow of support to Labour amongst overweighted respondents brought long standing problems of representativeness into the harsh light of day.

2 The 2015 British Election Study

This article uses two sources of data from the BES to examine what led to the polling errors: the BES Internet panel (Fieldhouse et al. 2015b) and the BES face-to-face post-election probability survey (Fieldhouse et al. 2015a).[3]

The 2015 BES Internet panel (BESIP) consists of six waves (at the time of writing) conducted between February 2014 and May 2015, with each wave interviewing around 30,000 respondents. The most important waves are the pre-election wave (wave 4) conducted between 4th and 30th March 2015; the campaign wave (wave 5) conducted as a rolling daily survey across the campaign period between 31st March and 6th May 2015; and the post-election wave (wave 6) conducted between 8th May and 28th May 2015. In total 26,112 respondents took all three election waves (a retention rate of 83% across waves 4, 5 and 6). Importantly, the BESIP is a non-probability sample (sampled from YouGov's opt-in pool of respondents), so representativeness has to be achieved through demographic quotas and weighting.

We also use the BES face-to-face survey: an address-based random probability sample of eligible voters living in 600 wards in 300 Parliamentary Constituencies in England, Scotland, and Wales. 2,987 people completed the face-to-face survey. Fieldwork was conducted by GfK between May 8th 2015 and September 13th 2015, achieving a 55.9% response rate.[4]

The BESIP is only one data source and other polls may have separate methodological issues we do not address directly here. However, the BESIP is well suited for studying polling errors in several ways. It is panel data, which allows us to compare what people intended to do before the election with what they did afterwards. It also has a substantial sample size, allowing us to examine issues such as weighting in a more disaggregated way by looking at smaller subgroups. There are also many variables available to examine different hypotheses around the polling errors, as the BES surveys ask a much broader range of questions than typical political polls. The dataset also has continuity with BES studies run in previous elections, particularly 2010, so the analysis of the 2015 polls can be

[3] Data are available through http://britishelectionstudy.com/
[4] Using response rate 3 (RR3) (AAPOR 2008).

compared with previous elections. Importantly, the BESIP is conducted by YouGov – one of the most prolific UK pollsters –meaning that the data is directly relevant to problems with the polls.

The BES face-to-face survey is an address-based random probability sample. Random probability samples are a gold standard with which to compare the data collected using other polling methods and are particularly useful for addressing questions about sample representativeness. Finally, unlike proprietary political polls, the BES surveys are released as open datasets so any analyses in this article can be checked and elaborated on by other researchers.

3 Evidence for polling error explanations

This section outlines the evidence in favour or against each of the above hypotheses.

3.1 Late swing

Figure 1 shows a possible upswing in voters intending to vote Conservative on the last 2 days of the campaign wave in the BES data. This data is based on the vote intentions of BES respondents interviewed on each day of the election campaign.

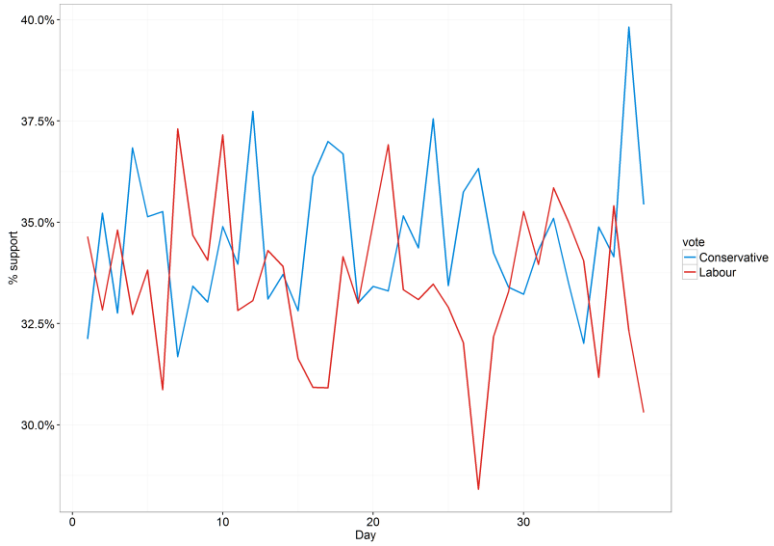

Figure 1. Vote intention by day in the British Election Study campaign wave. Figures do not filter by turnout likelihood and include squeezed responses.

While this data is consistent with late swing, it is not conclusive. With only around 1,000 respondents each day, these estimates have substantial sampling error. In addition, it is possible that supporters of one party may be more willing to take the survey on some days (e.g. when their party has a good day in the campaign) than others. This effect can exaggerate apparent changes in support for a party (Gelman et al. 2016).

When we look at the full sample in the post-election re-contact survey, we find no evidence of a late swing. Figure 2 shows the average vote intention in the campaign and post-election waves, finding essentially no difference in the average lead for the Conservatives.

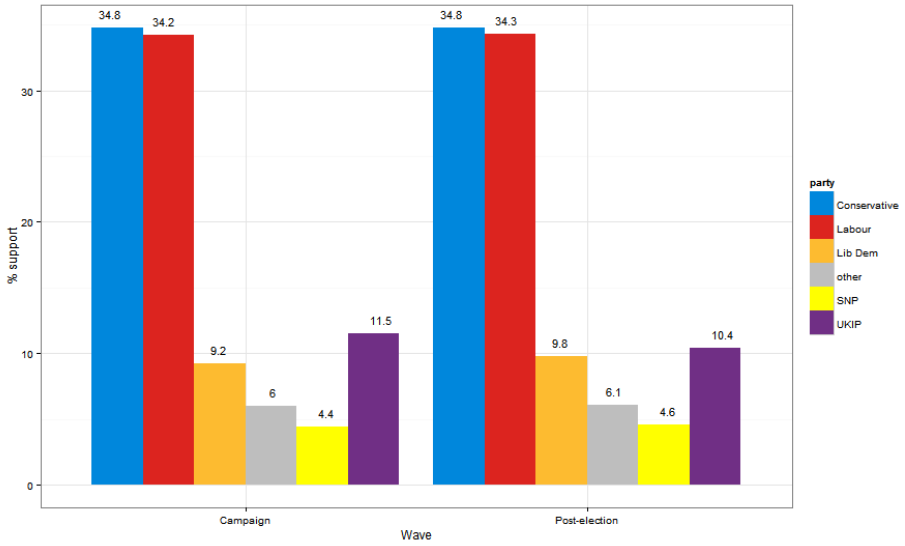

Figure 2. Average vote intention across the British Election Study campaign wave and post-election wave.

3.2 Differential turnout

Looking at self-reported turnout among respondents who report supporting different parties during the campaign shows little evidence of a differential turnout effect for Labour and the Conservatives. Those that expressed a party preference during the campaign but said they did not vote when asked after the election were slightly Labour leaning during the campaign, with these respondents having a 1.2% Labour lead over the Conservatives. Eventual non-voters make up only a small proportion of party supporters in the campaign (3.7% of Labour supporters and 3.5% of Conservative supporters during the campaign said they did not vote in the post-election wave).

However, differential turnout is difficult to detect because the same people who overstate their likelihood of voting before an election may also retrospectively misreport their turnout. To check whether turnout misreporting affects our analysis of differential turnout, we used the validated-voter subset of the BESIP. In total, we checked 2,700 internet panel respondents against the marked register. Of these respondents, 2,453 were assigned a good turnout estimate. The remainder had insufficient information to confirm whether their absence from the electoral register was due to non-registration or missing information about the respondent and/or the electoral register.

In total, 86.9% of these respondents report voting in the 2015 election. However, the validation exercise suggests that the true figure is likely to be around 76.0%, meaning that around 10.9% of the total sample reported voting but appear not to actually have voted. For more details on the data coding exercise, the coding decisions, reliability estimates and full tables of differential turnout for all parties see appendix B.

As Figure 3 shows, validated turnout rates among respondents who stated they would vote Conservative in the campaign wave are 1.3 percentage points higher than the turnout rates for

Labour.[5] However, this gap is not statistically significant. Since the sample size is relatively small, the evidence is consistent with either a small or zero differential turnout effect. In addition, we should bear in mind that most pollsters already filter for turnout likelihood (although YouGov are an exception[6]). Given the small effect of differential turnout, it is likely that existing turnout filtering is already sufficiently accounting for differential turnout.

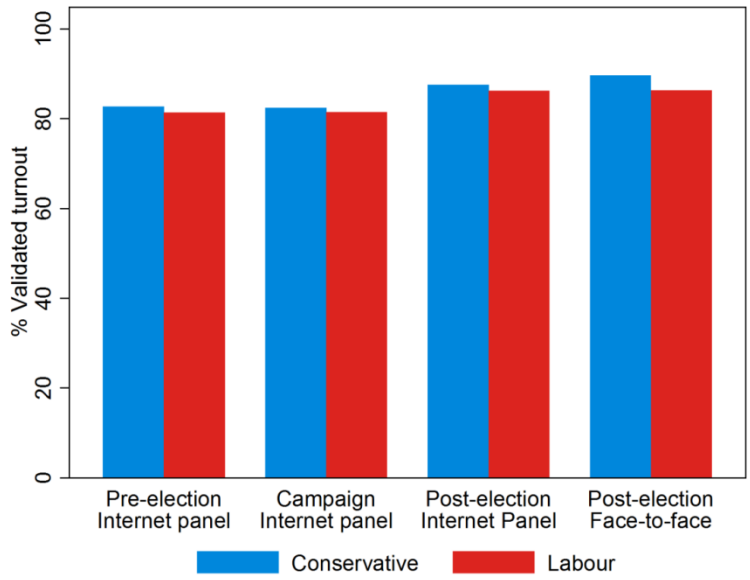

Figure 3. Validated turnout rates for self-identified voters of each party in pre-election, campaign and post-election BESIP waves and post-election face-to-face survey. None of these comparisons show a statistically significant difference in Conservative and Labour turnout rates.

We can also examine differential turnout misreporting after the fact by looking at the probability of accurate turnout reporting among self-reported voters of different parties in the BES face-to-face (shown in the last two groups of Figure 3). Both the face-to-face and panel show a statistically non-significant lower tendency of self-reported Labour voters to turn out to vote. Even if the difference were significant, it would not account for a major part of the polling miss at least in the BES Internet panel. There are some indications of differential turnout among the other parties (for instance lower turnout among UKIP respondents), but the sample size makes it difficult to draw strong conclusions about these differences. In any case, minor-party differences are unlikely to have caused the incorrect Conservative-Labour lead in the polls.

3.3 Differential "don't knows"/Spiral of Silence

In the campaign wave sample, 6.7% of people of people said that they "don't know" who they would vote for. In the post-election survey (when we can see how undecided respondents ended up voting), we find a small lead for the Conservatives among previously undecided voters: 30% said they voted

[5] Note that all comparisons here are shown using unweighted figures, as later sections of this paper indicate that there are substantial problems with the standard YouGov weights. The substantive results are not affected by the use of weights.
[6] http://ukpollingreport.co.uk/faq-turnout

Conservative and 28% said they voted Labour. However, this difference would only contribute around 0.05 percentage points towards the polling gap.

It is difficult to assess the spiral of silence and differential response to surveys directly as, by definition, respondents who are reluctant to reveal their political preferences would not be in the survey. We can however use the BESIP to look for indirect evidence of differential survey response. If Conservative voters are less likely to respond to opinion polls in general but at least some of them respond to a particular survey, we would expect, *ceteris paribus*, that they would be less likely to respond to future surveys. We can examine differential retention by looking at wave-on-wave response rates in the BESIP. Figure 4 shows the wave-on-wave retention of wave 1 BESIP respondents by their vote intention in wave 1 and their 2010 vote choice (recorded close to the 2010 election). Indeed Figure 4 shows no evidence supporting the spiral of silence theory and in fact shows a slight Conservative bias in survey response rate.

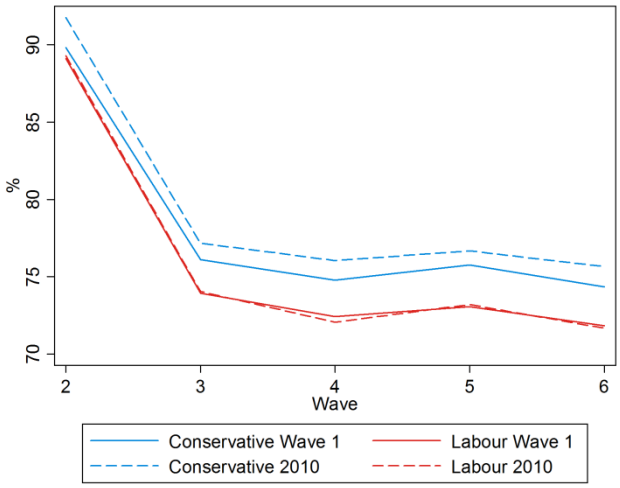

Figure 4. BESIP wave-on-wave retention for respondents who took the first wave of the study by vote choice in wave 1 (February 2014) and 2010 vote.

3.4 Vote intention misreporting

Vote intention misreporting is difficult to assess because there is no way to externally validate what respondents' actual vote choice is. However, we can look for indirect evidence of vote intention misreporting and do so in three ways: testing question ordering effects, examining the geographic distribution of polling error, and imputing vote choice for respondents who refuse to answer the vote choice question.

Several pollsters have suggested that there is a question ordering effect, where there is a higher rate of vote intention misreporting when vote intention is put before other questions in the survey. The BESIP gives some insight into this theory. In the first three waves, the placement of the vote intention question was randomized to be either at the start of the survey or at the end after all the other questions.[7] Figure 5 shows the results of this experiment: we find no significant differences in the

[7] The party identification question was separately randomized so that it would appear at the start or end of the survey as well in order to test whether there were differences in placing the questions together or apart.

proportion of respondents intending to vote for the Conservatives depending on whether the vote intention question was placed at the start of end of the survey (the same results hold when looking at Labour vote).[8]

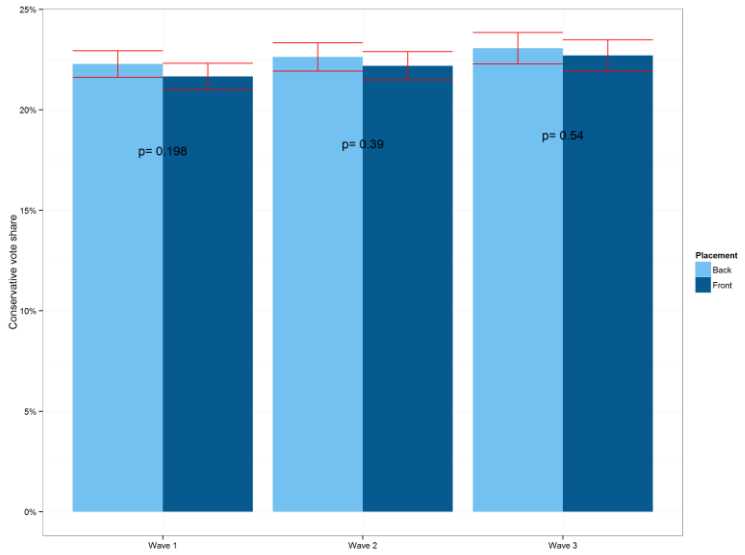

Figure 5. Conservative vote share by experimental group (note that the same randomization was maintained across waves, so the p values are not independent) P values are calculated using chi squared tests.

Another possible explanation would be that face-to-face surveys elicit fewer inaccurate vote choice responses than Internet panels. This explanation seems extremely unlikely, given that existing evidence points to a smaller social desirability effect online (where there is greater anonymity) than in person (Kreuter, Presser, and Tourangeau 2008; Dodou and de Winter 2014).

We can also examine the vote intention misreporting theory by considering where there is likely to be social pressure on Conservative voters. For example, it seems unlikely that Conservative voters would need to be shy in the heavily Conservative Shires but it is more plausible that they would be by shy in traditional Labour heartlands such as Sunderland.

Figure 6 shows that we actually observe the opposite pattern. The deviation between the proportion of BESIP respondents in a constituency saying they voted Conservative and the actual proportion of voters who voted Conservative is highest in strong Conservative areas, where we would expect the least social pressure against voting Conservative. By contrast, the same analysis on the BES face-to-face data again suggests that the primary problem with 2015 polls is representativeness. There is a close geographic match between the level of Conservative voting in the face-to-face survey and the actual election results. In other words, there are broadly the right number of Conservatives in the right places and the pattern is the opposite of that in the Internet panel.

[8] The sample sizes for this experiment are substantial, 30,192, 26,854, and 22,350 respondents, so the experiments have more than enough power to detect effects of the size suggested.

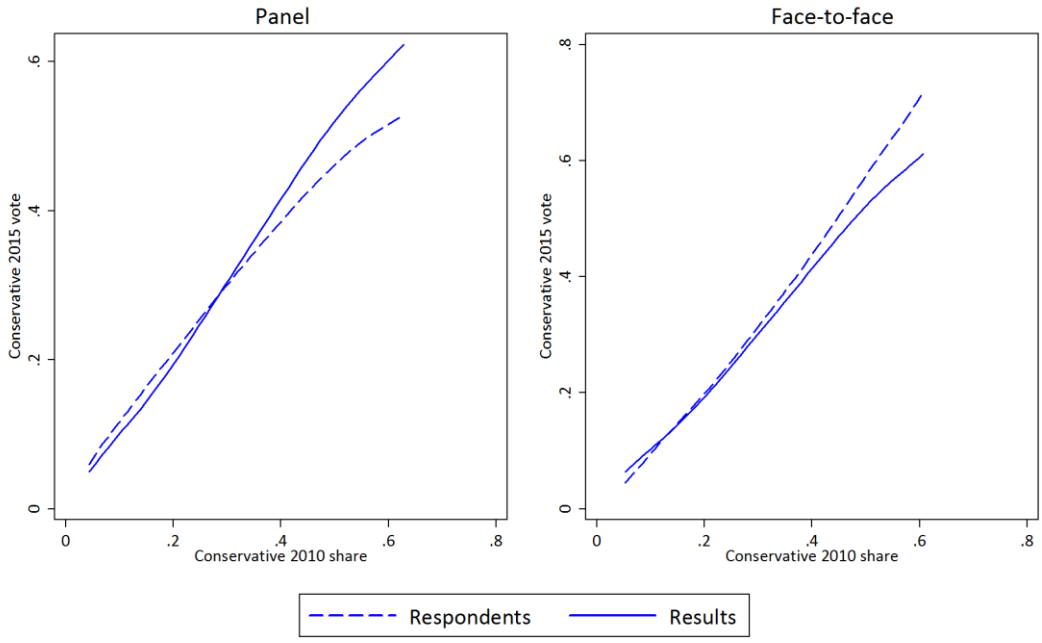

Figure 6. 2015 Conservative vote share among survey respondents (BESIP and BES face-to-face) and actual results according to 2010 Conservative (local polynomial smooth, bandwidth = 0.1).

We can also look more directly at those respondents who chose not to reveal their vote choice. In the BES face-to-face survey 4% of respondents refused to say who they voted for. After the 1992 polling disaster, differential party support in post-election surveys amongst those who had refused to reveal a party preference in pre-election surveys was suggested to be a contributing factor to the 1992 polling miss (Jowell et al. 1993b). The 2015 BES face-to-face survey is a cross-sectional survey so we cannot examine pre-election refusals. We can, however, use answers to a set of questions that asks 'How likely is it that you would ever vote for each of the following parties?' to estimate the likely proportions of party support amongst post-election refusals.[9] We estimate a conditional logit model of vote choice using respondents who revealed their vote choice to estimate the predicted probability of voting for each party based on how likely respondents said they were to ever vote for each party, controlling for party specific effects using party fixed effects. This simple model fits the data very well, correctly predicting 84% of party choices. Table 1 shows the expected proportion of refusals voting for each party, estimated by fitting the model to respondents who declined to reveal their vote choice. This analysis shows that, on the whole, refusals are spread between parties in a similar distribution to those respondents who reveal their vote choice. The expected proportion of refusals voting Conservative is actually smaller than the proportion who revealed their vote choice, and the difference for the Conservatives is larger than for any other party, suggesting that Conservative voters were the *least* shy in the survey.

[9] The BESIP did not have a "refuse" option in the vote choice question. The only "refusal" option would be for the respondent to say they did not know who they would vote for. As we show earlier in the paper there is no evidence to suggest that the pre-election "don't knows" voted in favour of the Conservatives.

Party	Respondents	Expected proportion amongst refusals	Difference
Conservative	39.87	34.2	-5.67
Labour	32.28	31.88	-0.4
Liberal Democrat	7.53	11.06	3.53
UKIP	11.16	13.13	1.97
Green Party	3.15	5.91	2.76

Table 1. Estimated percentage of refusals supporting each party in the BES face-to-face survey.

3.5 Representativeness/weighting

Pollsters try to overcome any deficiencies in their sampling process by weighting respondents to known targets. However there are many ways that samples may be unrepresentative of the population even after the weighting targets are reached. Section 3.2 suggests that BESIP respondents are more politically engaged than the general population (due to their higher rates of turnout).

The most important test of whether polling samples are representative is to compare the results of the BESIP (which, like political polls, uses weighting to achieve representativeness), to the BES face-to-face probability sample. The probability sample received a 56% response rate, making its representativeness much less assumption based than Internet panels and low response rate phone polls.

	GB	Face-to-face		BESIP	
	Result	Post-result	*Error*	Post-result	*Error*
Conservative	37.67	40.53	*2.86*	34.76	*-2.91*
Labour	31.16	32.75	*1.59*	34.32	*3.16*
Liberal Democrat	8.05	7.05	*-1.00*	9.81	*1.76*
SNP	4.85	4.69	*-0.16*	4.58	*-0.27*
Plaid Cymru	0.61	0.48	*-0.13*	0.54	*-0.07*
UKIP	12.88	10.61	*-2.27*	10.42	*-2.46*
Green Party	3.84	3.15	*-0.69*	4.31	*0.47*
BNP	0.01	0.16	*0.15*	0.07	*0.06*
Other	0.95	0.58	*-0.37*	1.19	*0.24*
Total			*8.70*		*11.39*
Con-Lab lead	6.51	7.78	*1.27*	0.44	*6.07*

Table 2. Performance of face-to-face and BESIP post-election surveys compared to GB results. Face-to-face weighting is based on NOP-provided weights accounting for region, age and gender and panel weighting uses the publicly available YouGov-provided regional weights.

Table 2 clearly shows that the BESIP has substantial problems measuring vote choice in the post-election wave. By contrast, we show that the face-to-face survey comes much closer to the correct result and has almost exactly the correct Conservative-Labour lead. In fact, unlike the polls, the face-to-face survey actually has too many Conservative voters. It is worth considering other possible explanations for the gap in the result between the face-to-face and BESIP survey. The face-to-face survey had a considerably longer field period (129 days in order to maximise response rate) compared to the BESIP (conducted over the course of 19 days). It is therefore possible that respondents started

mis-recalling their vote choices towards the end of this period. This phenomenon is well documented over longer periods (Durand, Deslauriers, and Valois 2015; Weir 1975; Himmelweit, Biberian, and Stockdale 1978; Elsas et al. 2013), so could be operating here. Appendix A analyzes this possibility and concludes that any trends in vote intention are driven primarily by different fieldwork speed in different constituencies.

Given the gap between the face-to-face data and the BESIP, and the lack of alternative explanations for it, it seems extremely likely that unrepresentative samples are a leading cause of the polling miss, at least in the BESIP.

3.5.1 How do the Internet and face-to-face samples differ?

The first difference between the face-to-face survey and the online panel is the level of political engagement. Figure 7 shows three examples of this difference: online panellists are much more interested in the campaign, more than twice as likely to be party members, and are much more likely to report having voted. In fact, 91.2% of the panel reported voting in the 2015 election, compared with 73.3% of face-to-face respondents. This turnout gap has substantial effects on sample representativeness because polling samples in the UK are weighted (and often quota sampled as well) to match population characteristics rather than voter characteristics.

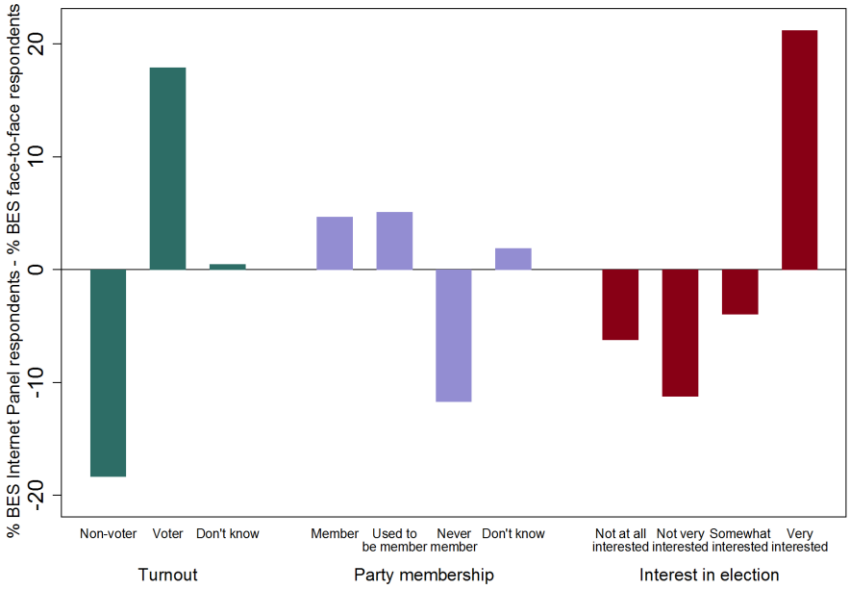

Figure 7. Comparisons of political engagement variables between the BESIP and face-to-face surveys. Positive (negative) bars indicate that compared to the BES face-to-face, respondents in each of the categories are over (under) represented in the BES Internet panel.

This problem becomes clear when we look at the match between various demographic variables across the face-to-face survey and the panel. In five out of six variables examined (see appendix C), there is a closer match between the two surveys when looking at the full sample than there is when looking at only the voters in each survey. In particular, the party identification distribution in Figure 8 is similar between the surveys when looking at the full population but greatly underestimates the Conservatives when looking only at the voters in each survey. The difference between the full and voter only samples for each survey is even more starkly demonstrated by the distribution of age groups, shown in Figure 9. The age distributions for the full samples are almost the same, with very similar proportions of respondents falling into each age group, but huge differences in the age

distribution of the voter samples. For the voter samples, the BESIP greatly under-represents respondents over age 56, and over-represents respondents from the youngest two age groups.

In other words, the BESIP appears to be relatively representative of the population (in terms of these variables) but the voters in the BESIP are unrepresentative of the voting population. This may explain why the some pollsters have observed that unweighted data performs better than weighted data (Sanders et al. 2004; Twyman 2008).

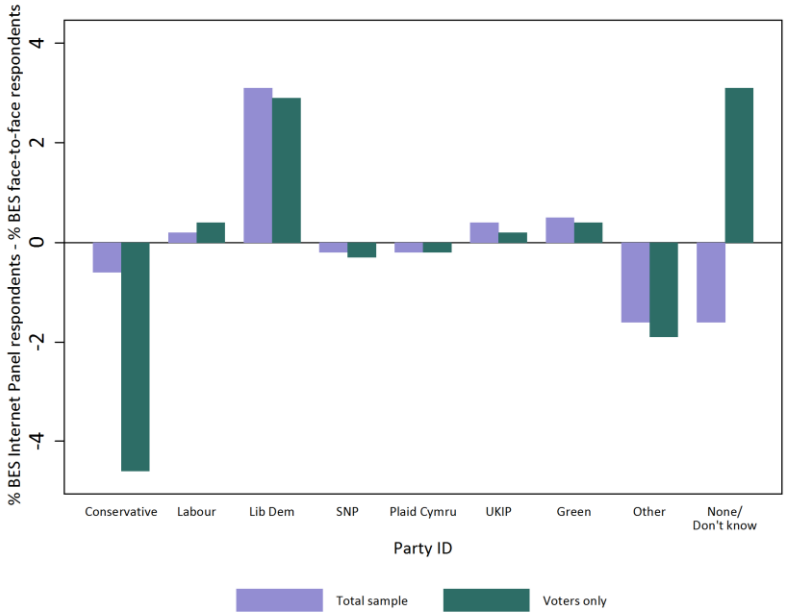

Figure 8. Comparison of party identification in face-to-face and BESIP surveys for all respondents and voters. Positive (negative) bars indicate that compared to the BES face-to-face, respondents in each of the categories are over (under) represented in the BESIP.

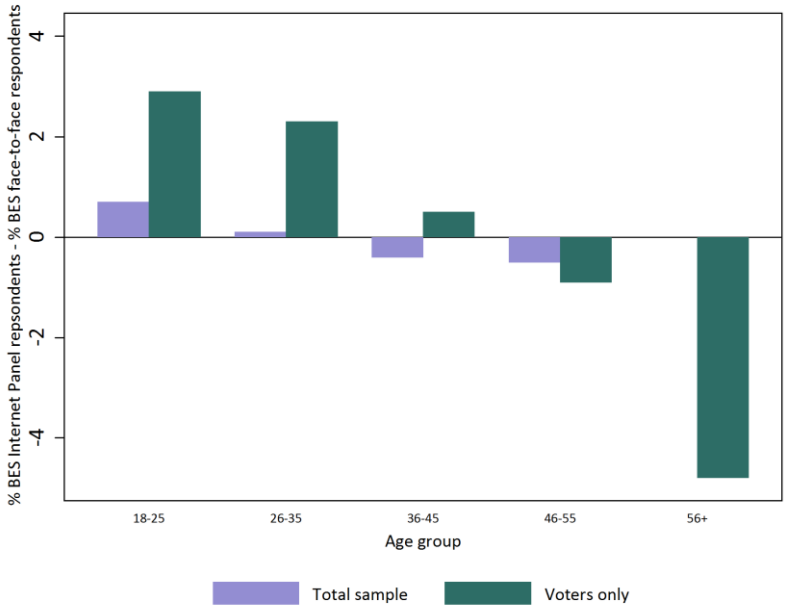

Figure 9. Comparison of age groups in face-to-face and BESIP surveys for all respondents and voters. Positive (negative) bars indicate that compared to the BES face-to-face, respondents in each of the categories are over (under) represented in the BESIP.

Another way to demonstrate the effect that standard sampling and weighting techniques have on the results is to look only at easily contacted respondents (i.e. the 714 respondents who responded on the first contact attempt). If this was the total sample, the response rate would have been around 13.5%. This response rate is still somewhat better than phone polls usually achieve, but nowhere near as high as the BES face-to-face achieved.

Table 3 shows the vote choices of respondents who answered on first contact and the full sample. The first two columns show the respondents' raw vote choices, while the second half of the table shows what happens when we weight each of the two samples to be representative of the population. An initial examination of the unweighted results suggests little difference in the partisan composition of people contacted on the first call and respondents contacted after several calls. The Conservative lead over Labour is larger in the full sample, but this difference is minor. However, reweighting the first contacts to represent the population in terms of region, age, and gender has a major effect on the Conservative party's lead over Labour in the sample. Whereas, the full sample barely changes after applying weights (the Conservative lead actually gets closer to the correct figure), the weighted first contact sample ends up with a Conservative lead of less than half the unweighted lead. This analysis again suggests that weighting unrepresentative samples to population values can have serious negative results.

	Unweighted		Weighted	
	Only first contacts (n=714)	All contacts (n=2955)	Only first contacts (n=714)	All contacts (n=2955)
Conservative	38.94	39.81	38.89	39.77
Labour	31.57	32.21	35.22	32.87
Liberal Democrats	6.62	7.55	4.8	7.17
SNP	624	4.86	3.79	4.65
Plaid Cymru	0.38	0.48	0.8	0.49
UKIP	13.42	11.25	14.11	11.01
Greens	2.08	3.12	1.78	3.28
BNP	0.19	0.14	0.21	0.18
Other	0.57	0.58	0.41	0.59
Con-Lab lead	7.37	7.6	3.67	6.9

Table 3. Reported vote choice among BES face-to-face respondents who responded on first contact and the final sample showing unweighted and weighted (using only the relevant subsample) results.

We can look at why this large shift happens by examining turnout across the weighting factors in the initial and final samples. We show this comparison for region in Table 4. The results show that self-reported turnout is higher in the initial sample in every region except for the East Midlands. However, the sizes of the differences vary substantially by region. The regions that most understate turnout in

the initial sample are Wales, Yorkshire, the North East, Scotland, Wales, and the North West, which are also the regions that the Conservatives perform worst in. In other words, voters are most incorrectly upweighted in the Conservatives' worst performing regions, meaning voters in these regions are upweighted and take the place in the sample of non-voters in these regions.

Region	Turnout			Conservative vote share 2015
	First contact	Final sample	Difference	
East Midlands	77.1	82.5	-5.4	43.5
East of England	72.6	71.2	1.4	49.0
London	73.1	71.2	1.9	34.9
West Midlands	72.7	70.3	2.4	41.8
South East	81.9	77.9	4	51.6
South West	86.3	81.6	4.7	46.5
North West	78.4	73.2	5.3	31.2
Scotland	82.5	77.2	5.3	14.9
North East	70	63.4	6.6	25.3
Yorkshire and the Humber	75.7	67.9	7.8	32.6
Wales	88	67.7	20.3	27.2

Table 4. Regional differences in turnout between BES face-to-face respondents who responded on first contact and the final sample (unweighted)

There are also differences in self-reported turnout rates for men and women among initial contacts and the final sample. Men who were initially contacted report turnout rates 7.6 percentage points higher than women. However, the final sample closes this gap to 0.7 percentage points. This gender gap also may have affected the results, as women have a considerably larger Conservative-Labour lead than men (9.3 percentage points versus 6.5 percentage points). The larger lead among women seems to be primarily due to stronger male support for UKIP and the SNP, reducing the male shares for Labour and Conservatives.

The picture is less clear in terms of age groups. The 18-24 age group report lower turnout in initial contacts than in subsequent contacts (44.2% versus 50%). However, the 25-39 age group report higher turnout (63% versus 57.7%) and the 60+ age group have almost no gap (87.3% versus 86.4%). In a linear regression, the relationship between age and turnout increases slightly in the full sample compared with the first contact sample.

Overall, the initially contacted sample would have led to a far worse result than the final sample, after weighting, due to the early interviews disproportionately contacting voters, as opposed to non-voters,

in more Conservative leaning regions. The sample would also have had too many male voters (who showed a smaller Conservative-Labour lead).

One way to tackle this problem for post-election surveys is to add in an additional weighting factor of the expected proportion of non-voters. This adjustment means that non-voting respondents are upweighted to 40% of the sample (we use voting age population turnout), while the other weighting targets are kept constant. The consequence of including turnout as a weighting factor is that respondents from demographics that vote less will be upweighted in the non-voting subsample, and respondents from low turnout demographics will be correspondingly downweighted in the voting subsample.

Table 5 demonstrates this adjustment applied to the post-election wave of the BES internet panel. The raked with non-voting column shows that the performance of the post-election wave greatly increases, with the sum of absolute errors decreasing by 3.2 percentage points and the Conservative lead increasing by 2.3 percentage points. The performance of the BESIP increases further once we weight party members to the correct proportion of the population. In fact, the performance of the post-election wave in absolute error terms is better than the face-to-face survey. This adjustment still leaves a substantial portion of the Conservative-Labour lead to explain, but is clearly an improvement over implicitly assuming that the stated turnout of respondents is accurate.

We can also look at this problem in the other direction, by applying weights to the face-to-face data and seeing whether it can be made to exhibit the same problems as polling data. The Conservative lead over Labour in the face-to-face survey falls by more than half when we reweight the face-to-face using only the respondents who reported voting.

	GB	BES Face-to-face		Post-election BESIP		
	Result	Raked full	Raked only voters	Raked	Raked with non-voting	Raked/non-vote party member
Conservative	37.67	39.95	38.39	34.40	36.13	36.44
Labour	31.16	33.04	34.96	33.48	32.94	32.75
Liberal Democrat	8.05	7.18	6.95	9.37	9.17	9.15
SNP	4.85	4.53	4.38	3.98	4.47	4.34
Plaid Cymru	0.61	0.46	0.49	0.49	0.49	0.49
UKIP	12.88	10.88	10.69	11.99	11.38	11.48
Green Party	3.84	3.21	3.37	4.90	4.22	4.13
BNP	0.01	0.19	0.20	0.08	0.07	0.08
Other	0.95	0.56	0.56	1.31	1.13	1.14
Sum absolute error	0	8.16	9.45	10.28	7.06	6.50
Con-Lab lead	6.51	7.23	3.43	0.92	3.19	3.70

Table 5. Face-to-face and panel data weighted to population and electorate. Note that weights are calculated for the BESIP post-election wave at the GB level and use separate age and gender targets. Face-to-face results compare raking age/gender and region targets on the full sample and voter only samples. The face-to-face weights are not exactly the same as those provided by NOP as the weights use the YouGov age bands.

4 Why 2015

Given that the main problem we identified with the polls in 2015 relates to the fundamental issue that polls are unrepresentative in terms of political attention, we must ask an important question: why did the UK polling industry perform so badly in 2015 rather than in some other election?

To help answer this question, we look at the initial weights/turnout adjusted weights for Labour, the Conservatives, and the Liberal Democrats in the post-election wave in Figure 10. As our reweighting results suggested, Labour voters are most heavily down-weighted in this adjustment and Conservative voters are least down-weighted. This tells us that Labour leaning respondents were most over-represented as voters in the sample in 2015.

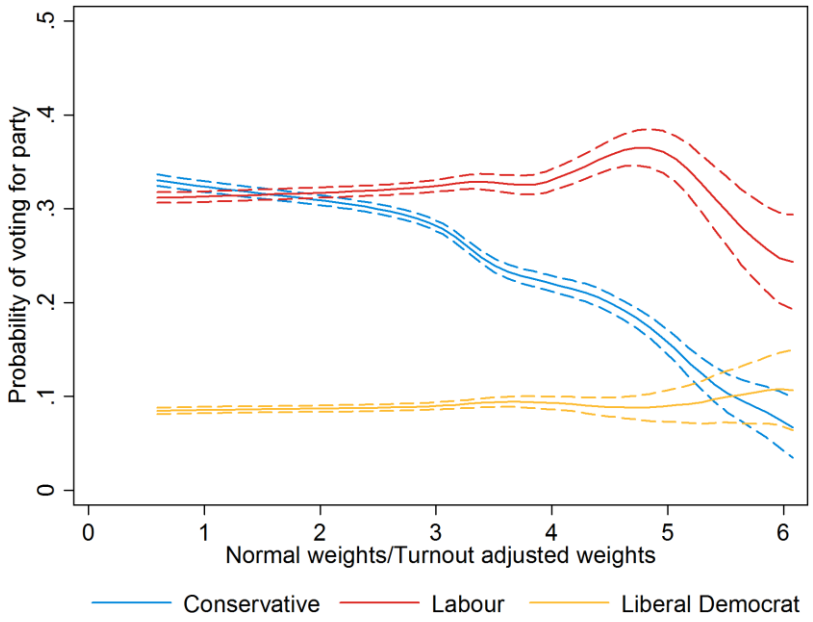

Figure 10. Probability of voting for a party in 2015 by the ratio of the normal and turnout adjusted 2015 weights.

We can also conduct the same exercise for the 2010 BESIP. Figure 11 shows these results for the 2010 data. Whereas there is essentially no relationship between down-weighting and Liberal Democrat vote in 2015, in 2010 the relationship is strong and positive. Conversely, the relationship between Labour vote share and adjustment is much weaker in 2010 than it was in 2015.

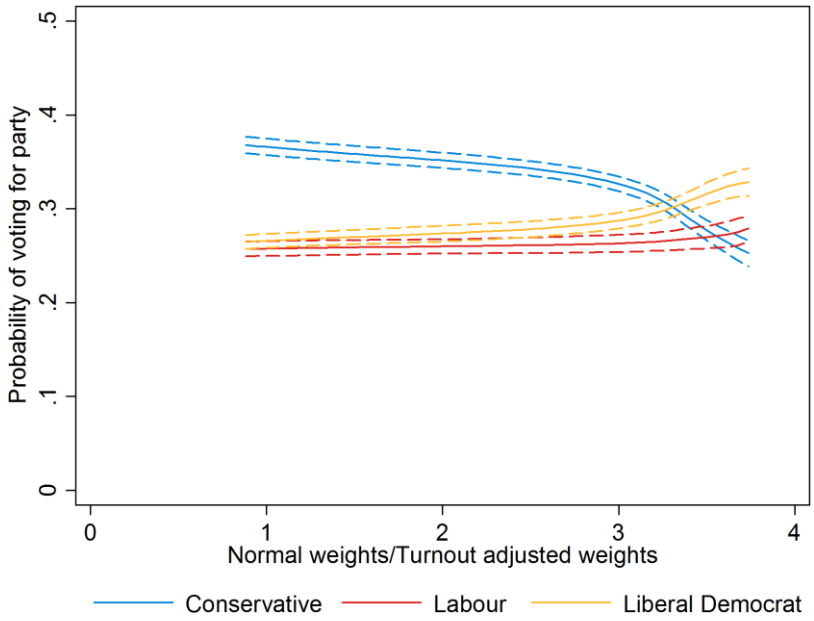

Figure 11. Probability of voting for a party in 2010 by the ratio of the normal and turnout adjusted 2010 weights

These results potentially explain the 2010 over-estimation of the Liberal Democrats (Pickup et al. 2011) Pickup et al. estimate that the 2010 polls over-estimated the Liberal Democrats by around 4.0 percentage points: a figure not that much lower than the 2015 polling miss. At the time they were unable to find a clear explanation for this over-estimate, but concluded that it was highly unlikely to be solely the result of a late swing away from the Liberal Democrats. Instead our findings suggest that this over-estimate results from polls over-sampling/overweighting voters from low turnout demographics and that these voters tend to support the Liberal Democrats. Using the turnout adjusted weights in 2010 reduces the Liberal Democrat's 2010 share by 0.9 percentage points, increases the Conservative share by 1.3 percentage points and increases the Labour share by 0.2 percentage points. We cannot adjust the 2010 BESIP in exactly the same way as we adjusted the 2015 BESIP because the 2010 BESIP does not include questions on party membership. It is therefore possible that the true effect of this mechanism is larger in 2010 than we find here.

Given these findings, we think that the most likely explanation for "why 2015?" is that the over-weighted respondents did not vote Labour in 2010. Given that the polls also over-estimated the Labour share of the vote in 2001 and 2005, it is possible that 2010 is actually the unusual election. The surprise surge of support for the Liberal Democrats among low-turnout demographic groups such as students may in fact have saved the pollsters from a 2015 polling miss in 2010.

5 Conclusions

BES data provides substantial evidence indications about what caused the 2015 polling errors.

The evidence for late swing is limited. While the campaign wave suggests some movement, the post-election wave doesn't reflect late swing. Similarly, our analysis suggests that respondents who said "don't know" during the campaign did not systematically shift in a way that changed the result.

We also think that vote intention misreporting is unlikely to have caused the polling error. We show strong evidence against question ordering affecting respondents' truthfulness about supporting the Conservatives. Additionally, the polling errors show the opposite geographic pattern than we would expect if social pressure led respondents to lie about voting for the Conservatives. Finally, refusals in the face-to-face support the Conservatives less than the rest of the sample.

Despite initial circumstantial evidence that differential turnout contributed to the polling miss, the validated vote data provides little evidence that differential turnout was a major cause of the polling miss. Our analysis shows that the validated turnout of respondents who supported different parties during the campaign is almost the same, with only a small and non-significant difference between the turnout of Labour and Conservative supporters.

By contrast, sampling and weighting played a substantial role in the polling miss. The face-to-face probability sample shows a much more accurate picture than the polling data and does not underestimate the Conservative vote share. Our subsequent analysis shows that the higher accuracy of the face-to-face compared to polling data is due to pollsters surveying too few non-voters, but still treating their samples as if they are population samples. Undersampling non-voters makes the voter subsample look too much like the general population (over-representing non-voting demographics).

There is still more to work out however. Our analysis identifies a likely cause of the miss (weighting to population targets without correcting for turnout), but more work is needed on how pollsters can correct this problem. Our article suggests one such approach that could be used for post-election samples, but more work is needed to determine what corrections pollsters can make in pre-election polls.

Finally, this article's findings need to be considered by pollsters in the UK and in other countries. Weighting to population targets and then removing non-voters is by far the most dominant election polling approach, and any pollster that uses it risks the same kind of polling miss that Britain saw in 2015. Given the increasing number of polling misses throughout the world, we may already be seeing the results of this methodological approach.

References

AAPOR. 2008. “Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.” http://www.aapor.org/AAPORKentico/AAPOR_Main/media/MainSiteFiles/Standard_Definitions_07_08_Final.pdf.

Abrams, Mark. 1970. “The Opinion Polls and the 1970 British General Election.” *The Public Opinion Quarterly* 34 (3): 317–24.

Boon, Martin, and John Curtice. 2010. “General Election 2010: Did the Opinion Polls Flatter to Deceive?” *Research Live*. July 6. http://www.research-live.com/article/opinion/general-election-2010-did-the-opinion-polls-flatter-to-deceive/id/4003088.

British Polling Council. 2015. “General Election: 7 May 2015.” http://www.britishpollingcouncil.org/general-election-7-may-2015/.

Burden, Barry C. 2000. “Voter Turnout and the National Election Studies.” *Political Analysis* 8 (4): 389–98.

Chang, Linchiat, and Jon A. Krosnick. 2009. “National Surveys Via Rdd Telephone Interviewing Versus the Internet Comparing Sample Representativeness and Response Quality.” *Public Opinion Quarterly* 73 (4): 641–78. doi:10.1093/poq/nfp075.

Couper, Mick P., and Peter V. Miller. 2008. “Web Survey Methods Introduction.” *Public Opinion Quarterly* 72 (5): 831–35. doi:10.1093/poq/nfn066.

Cowley, Philip, and Dennis Kavanagh. 2016. *The British General Election of 2015*. Houndmills, Basingstoke, Hampshire: Palgrave MacMillan. http://www.palgrave.com/gb/book/9781137366108.

Crewe, Ivor. 1992. “A Nation of Liars? Opinion Polls and the 1992 Election.” *Parliamentary Affairs* 45 (4): 475–95.

Curtice, John, Nick Sparrow, and John Turner. 1997. “The Missing Tories in Opinion Polls: Silent, Forgetful or Lost?” *British Elections & Parties Review* 7 (1): 23–41. doi:10.1080/13689889708412987.

Curtin, Richard, Stanley Presser, and Eleanor Singer. 2005. “Changes in Telephone Survey Nonresponse over the Past Quarter Century.” *Public Opinion Quarterly* 69 (1): 87–98. doi:10.1093/poq/nfi002.

Dodou, D., and J. C. F. de Winter. 2014. “Social Desirability Is the Same in Offline, Online, and Paper Surveys: A Meta-Analysis.” *Computers in Human Behavior* 36 (July): 487–95. doi:10.1016/j.chb.2014.04.005.

Durand, Claire, Mélanie Deslauriers, and Isabelle Valois. 2015. “Should Recall of Previous Votes Be Used to Adjust Estimates of Voting Intention?” *Survey Insights: Methods from the Field*. doi:10.13094/SMIF-2015-00002.

Elsas, Erika J. van, Rozemarijn Lubbe, Tom W. G. van der Meer, and Wouter van der Brug. 2013. “Vote Recall: A Panel Study on the Mechanisms That Explain Vote Recall Inconsistency.” *International Journal of Public Opinion Research*, October, edt031. doi:10.1093/ijpor/edt031.

Fieldhouse, Edward, Jane Green, Geoffrey Evans, Hermann Schmitt, Cees Van der Eijk, Jonathan Mellon, and Christopher Prosser. 2015a. “British Election Study, 2015: Face-to-Face Survey [computer File].” doi:10.13140/RG.2.1.2603.9764/1.

———. 2015b. “British Election Study, 2015: Internet Panel Survey [computer File].” doi:10.13140/RG.2.1.3625.1604.

Fisher, Stephen D., and Michael Lewis-Beck. 2016. “Forecasting the 2015 British General Election: The 1992 Debacle All over Again?” *Electoral Studies*. doi:10.1016/j.electstud.2015.11.016.

Gelman, Andrew, Sharad Goel, Douglas Rivers, and David Rothschild. 2016. “The Mythical Swing Voter.” *Quarterly Journal of Political Science*.

Górecki, Maciej A. 2011. “Why Bother Lying When You Know So Few Care? Party Contact, Education and Over-Reporting Voter Turnout in Different Types of Elections.” *Scandinavian Political Studies* 34 (3): 250–67. doi:10.1111/j.1467-9477.2011.00271.x.

Green, Jane, Edward Fieldhouse, and Christopher Prosser. 2015. “How Coalition Government Leads to Minor Party Voting (in the Absence of Proportional Representation): A Case Study of the 2015 British General Election.” Working paper.

Green, Jane, Will Jennings, and Jonathan Mellon. 2015. “Shaping the Vote Decision and the Outcome? The British Conservative Party’s 2015 Campaign on the Economy.” Working paper.

Himmelweit, Hilde T., Marianne Jaeger Biberian, and Janet Stockdale. 1978. “Memory for Past Vote: Implications of a Study of Bias in Recall.” *British Journal of Political Science* 8 (03): 365–75. doi:10.1017/S0007123400001411.

Holbrook, Allyson L., and Jon A. Krosnick. 2010. “Social Desirability Bias in Voter Turnout Reports Tests Using the Item Count Technique.” *Public Opinion Quarterly* 74 (1): 37–67. doi:10.1093/poq/nfp065.

Ipsos MORI. 2015. “The Ipsos MORI Final Election Poll.” https://www.ipsos-mori.com/researchpublications/researcharchive/3572/The-Ipsos-MORI-Final-Election-poll.aspx.

Jowell, Roger, Barry Hedges, Peter Lynn, Graham Farrant, and Anthony Heath. 1993a. “Who Misled Whom? The Polls and the Voters in the 1992 British Election.” In . St Charles, Illinois.

———. 1993b. “The 1992 British Election: The Failure of the Polls.” *Public Opinion Quarterly* 57 (2): 238–63. doi:10.1086/269369.

Karp, Jeffrey A., and David Brockington. 2005. “Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries.” *Journal of Politics* 67 (3): 825–40. doi:10.1111/j.1468-2508.2005.00341.x.

Keeter, Scott, Courtney Kennedy, Michael Dimock, Jonathan Best, and Peyton Craighill. 2006. “Gauging the Impact of Growing Nonresponse on Estimates from a National RDD Telephone Survey.” *Public Opinion Quarterly* 70 (5): 759–79. doi:10.1093/poq/nfl035.

Kellner, Peter. 2015. “Who Won the UK General Election, How and Why?” presented at the Who won the UK general election, how, why, and with what implications for Britain and Europe?, Europe House, London, May 14.

Kreuter, Frauke, Stanley Presser, and Roger Tourangeau. 2008. “Social Desirability Bias in CATI, IVR, and Web Surveys The Effects of Mode and Question Sensitivity.” *Public Opinion Quarterly* 72 (5): 847–65. doi:10.1093/poq/nfn063.

Market Research Society Working Party. 1994. *The Opinion Polls and the 1992 General Election.* London: Market Research Society.

Noelle-Neumann, Elisabeth. 1993. *The Spiral of Silence: Public Opinion--Our Social Skin*. University of Chicago Press.

Pedley, Keiran. 2015. “Polling Matters - Ep. 23 Labour’s Election Polling W/ James Morris.” http://kpedley.podbean.com/e/polling-matters-ep-23-labours-election-polling-w-james-morris/.

Pew Research Center. 2016. “Can Likely Voter Models Be Improved? Evidence from the 2014 U.S. House Elections.” http://www.pewresearch.org/2016/01/07/can-likely-voter-models-be-improved/.

Pickup, Mark, J. Scott Matthews, Will Jennings, Robert Ford, and Stephen D. Fisher. 2011. “Why Did the Polls Overestimate Liberal Democrat Support? Sources of Polling Error in the 2010 British General Election.” *Journal of Elections, Public Opinion and Parties* 21 (2): 179–209. doi:10.1080/17457289.2011.563309.

Presser, Stanley, and Michael Traugott. 1992. “Little White Lies and Social Science Models Correlated Response Errors in a Panel Study of Voting.” *Public Opinion Quarterly* 56 (1): 77–86. doi:10.1086/269296.

Sanders, David, Harold D. Clarke, Marianne C. Stewart, Paul F. Whiteley, and Joe Twyman. 2004. “The 2001 British Election Study Internet Poll.” *Journal of Political Marketing* 3 (4): 29–55. doi:10.1300/J199v03n04_02.

Singh, Matt. 2015. “One Confirmed Cause of the Polling Pig’s Ear - Some People WERE Telling Porkies.” *Number Cruncher Politics*. http://www.ncpolitics.uk/2015/05/one-confirmed-cause-of-the-polling-pigs-ear-some-people-were-telling-porkies.html/.

Streb, Matthew J., Barbara Burrell, Brian Frederick, and Michael A. Genovese. 2008. “Social Desirability Effects and Support for a Female American President.” *Public Opinion Quarterly* 72 (1): 76–89. doi:10.1093/poq/nfm035.

Turner, John, and Nick Sparrow. 1997. “Hearing the Silence: The Spiral of Silence, Parties and the Media.” *Media, Culture & Society* 19 (1): 121–31. doi:10.1177/016344397019001009.
Twyman, Joe. 2008. “Getting It Right: YouGov and Online Survey Research in Britain.” *Journal of Elections, Public Opinion and Parties* 18 (4): 343–54. doi:10.1080/17457280802305169.
Weir, Blair T. 1975. “The Distortion of Voter Recall.” *American Journal of Political Science* 19 (1): 53–62. doi:10.2307/2110693.
Wilks, S. S. 1940. “Representative Sampling and Poll Reliability.” *Public Opinion Quarterly* 4 (2): 261–69. doi:10.1086/265400.
Worcester, Robert. 1996. “Political Polling: 95% Expertise and 5% Luck.” *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 159 (1): 5–20. doi:10.2307/2983464.

Appendices to

Missing Non-Voters and Misweighted Samples: Explaining the 2015 Great British Polling Miss

Appendix A- Time trends in face-to-face responses

The tendency for voters to misremember their past vote is well documented (Durand, Deslauriers, and Valois 2015; Weir 1975; Himmelweit, Biberian, and Stockdale 1978; van Elsas et al. 2013). One particular concern with the British Election Study face to face survey is that because the fieldwork period was quite long, those surveyed later on may be more likely to misremember their vote. If this misremembering took the form of a drift towards the winner of the election (the bandwagon effect) this would inflate the Conservative share of the vote. If this was the case, the apparent representativeness of the face to face survey may in fact be down to a different sort of error and so it would not be able to tell us much about the causes of the polling miss.

Figure A1 shows that when examined longitudinally there is an increasing Conservative lead in the face-to-face survey as the field work progressed. The blue line plots this lead against the respondents in the survey in the order they were surveyed and shows that after about 200 voters (this analysis excludes non-voters) were surveyed, the lead moves in a linear fashion from a 4% Labour lead over the Conservatives to an almost 8% lead of the Conservatives over Labour. At first glance this certainly appears to be problematic for the representativeness of the face-to-face survey.

However this trend, which appears to be driven by time on first examination, is actually an artefact of geography. Even if we had a perfectly representative survey and all respondents could remember their votes accurately, if different areas are not sampled simultaneously then the fact that different parties have different levels of support in different constituencies may give the appearance of longitudinal differences, even when the differences are geographic. The orange line in the graph below illustrates this point by showing what the Conservative-Labour lead looks like using the constituency results of the respondents, rather than their reported vote, by the order in which they were surveyed. Of course the face-to-face survey does not match this line exactly, but it is remarkably close and the longitudinal trend for an increasing Conservative lead is very similar in both lines.

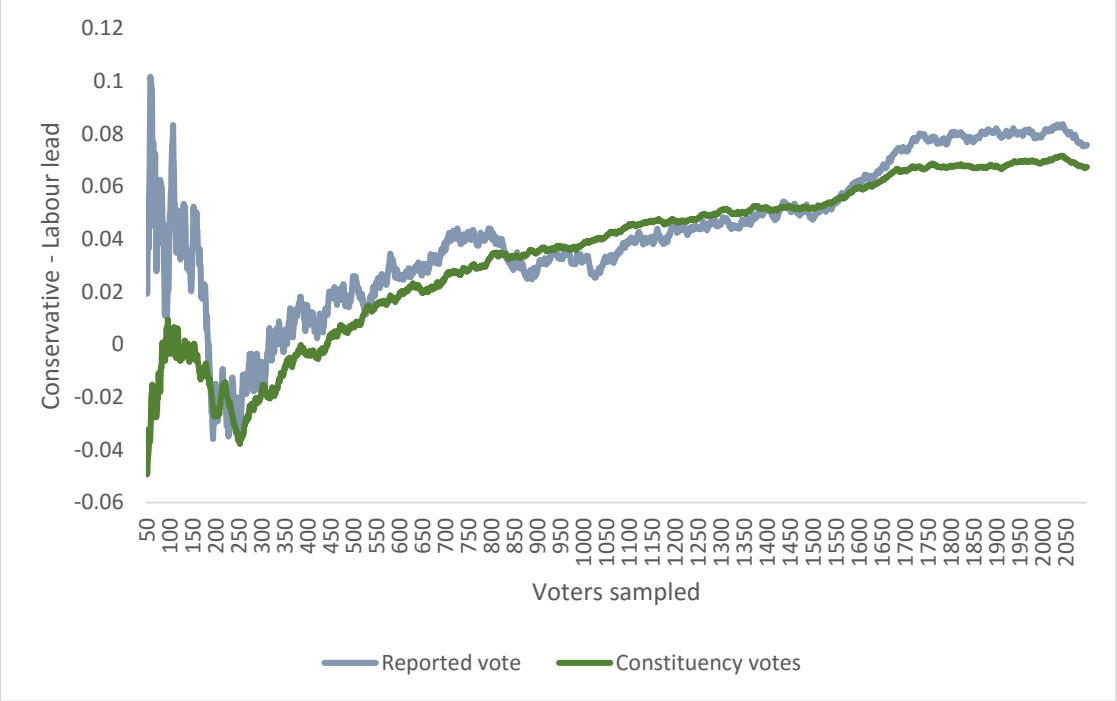

FIGURE A1.. CONSERVATIVE-LABOUR LEAD IN REPORTED AND CONSTITUENCY RESULTS BY THE ORDER RESPONDENTS WERE SAMPLED.

Figure A2 makes the same point using the Conservative and Labour shares of the vote. Although the face-to-face survey slightly oversamples Conservative and Labour voters, this is fairly consistent across time and does not appear to be a result of any 'bandwagon' effect. Simply from knowing the distribution of voters in different constituencies we would expect there to be an increasing proportion of Conservatives and a decreasing proportion of Labour voters for the order in which the respondents were sampled.

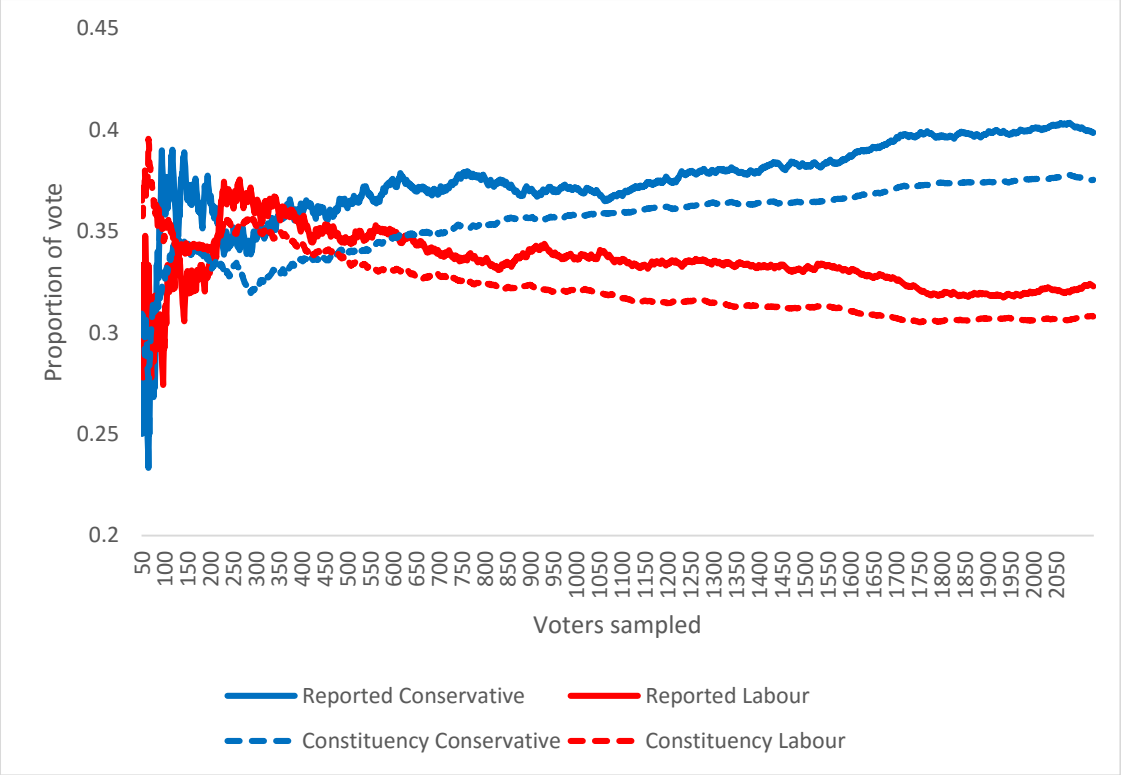

FIGURE A2.. CONSERVATIVE AND LABOUR VOTE SHARE IN REPORTED AND CONSTITUENCY RESULTS BY THE ORDER RESPONDENTS WERE SAMPLED.

Another way of examining the question of the impact of the long fieldwork period on the representativeness of the face-to-face survey is to look at reachability in general rather than at which groups are missing in particular. If the face-to-face survey is more representative because it is able to reach people who are more difficult to contact (because it repeatedly tried to contact those who it could not get hold of rather than issuing a second sample) then this could explain why polls, which are necessarily conducted in a much quicker fashion, were less representative. One method of looking at this is to compare the results of the face-to-face survey across respondents who answered after relatively few attempts, to the results from respondents who only answer after they are contacted many times. The theory is that easy to contact respondents most resemble samples that pollsters manage to contact (Jowell et al. 1993b). However, the analysis shows that we appear to observe the opposite pattern.

We combine the question of longitudinal differences and number of calls and look at both with a series of multivariate logistic models. Table A1 shows a logit model (model 1) and a multilevel logit model with observations clustered at the constituency level (models 2 and 3) for voting either Labour or Conservative. In model 1 the interview time has a statistically significant effect for both parties – respondents were less likely to say they voted Labour over time and more likely to say they voted Conservative. Moving to the multilevel analysis, simply adding a random intercept for constituency (model 2) decreases the effect of time for both parties slightly, and also shows that there is considerable variance in voting for each party at the constituency level (as you would expect). When the share of the vote that the party received in the constituency (model 3) the effect of time is much smaller and no longer statistically significant and the share of the vote accounts for almost all of the constituency level variation in the probability of voting for either party. Models 1-3 for the Labour shows that the probability of being a Labour voter increases with the number of calls that were needed to contact the respondent. Model 1 for the Conservatives shows that voters were less likely to be

Conservative the more calls that were needed to contact them, though this effect disappears once constituency variation is accounted for.

	Labour			Conservative		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Number of calls	0.105***	0.0905***	0.0614**	-0.0439**	-0.0318	-0.0236
	(0.0231)	(0.0283)	(0.0250)	(0.0222)	(0.0262)	(0.0236)
Interview time	-7.50e-11***	-7.27e-11***	-3.29e-11	6.56e-11***	5.29e-11**	2.20e-11
	(2.21e-11)	(2.78e-11)	(2.40e-11)	(2.02e-11)	(2.48e-11)	(2.18e-11)
Party vote share in constituency			0.0508***			0.0477***
			(0.00346)			(0.00351)
Constant	130.3***	126.1***	54.98	-115.1***	-92.89**	-40.66
	(38.69)	(48.65)	(41.93)	(35.40)	(43.38)	(38.16)
Constituency level variance		0.804***	0.00827		0.577***	0.00549
		(0.162)	(0.0663)		(0.129)	(0.0600)
N	1934	1934	1934	1934	1934	1934

TABLE A1. LOGISTIC REGRESSION MODELS OF THE RELATIONSHIP BETWEEN INTERVIEW TIME AND NUMBER OF CALLS ON RECALLED LABOUR AND CONSERVATIVE VOTE.

Overall there is no evidence from this analysis that the polls were capturing a group of respondents similar to those which would be obtained by a cursory probability sample that did not try to recontact respondents. Importantly, there's no great Labour lead among easy to reach voters. Perhaps this is not that surprising given that easy to reach voters are quite different across modes. However, this method only partially reflects reachability for pollsters. While reachability does not seem to be an important predictor in the face-to-face context, there may be a greater reachability bias among Internet samples and phone polls.

It is also worth noting is that political attention does not predict the number of calls, suggesting that failure to reach is less a function of attention than of difficulty contacting. Then again, we have survivor bias here. If non-responders are a mixed group of uninterested and busy, the eventual responders will be differentially from the busy group. However, if that was the case, then it would suggest that effort towards increasing response rates wouldn't reduce the political interest bias.

Appendix B - Validated vote

The name and address information of face-to-face respondents who had given their permission for their information to be linked with the electoral registers was matched against the marked electoral registers. Additionally, respondents from the internet panel who happened to live in the same wards that were sampled for the face-to-face survey were also matched against the registers.

Based on this information respondents were coded into five categories:

1) Voted: The respondents appeared on the electoral register and was marked as having voted.

2) Not voted - registered: The respondent appeared on the electoral register but was not marked as having voted.

3) Not voted - unregistered: The respondent did not appear on the electoral register but there was sufficient information to infer that they were not registered to vote, e.g. other people were registered to vote at the address or if no one was registered at the address people were registered at surrounding addresses.

4) Insufficient information: We did not have sufficient information in the registered to assess whether the respondent was registered and voted, either because we were missing the necessary pages from the register or we had not been sent the register.

5) Ineligible: the respondent was on the electoral register but was marked ineligible to vote in the general election.

	Face-to-face		Panel	
Vote validation status	Freq	Percent	Freq.	Percent
Voted	1,286	54.49	1,863	69
Not voted-registered	387	16.4	188	6.96
Not voted-unregistered	300	12.71	402	14.89
Insufficient information	370	15.68	228	8.44
Ineligible	17	0.72	19	0.7
Total	2,360		2,700	

TABLE B1. VALIDATION RESULTS FOR FACE-TO-FACE AND PANEL

Missing registers

Despite the heroic efforts of the Electoral Commission many Local Authorities did not supply their marked electoral registers, in all we were missing information from around 15% of the face-to-face respondents who agreed to be matched. Unfortunately the areas we are missing areas for a disproportionately Conservative (based on the 2015 Election constituency results) which introduces a slight Labour bias to the reported vote amongst those who had their vote validated. There is a larger Labour bias among the validated respondents from the internet panel. We are still assessing the extent to which this is driven by the same problem of missing registers or whether it is due to the clustering of panel respondents in our convenience sub-sample of the panel.

Reliability

Based on multiple coding of a subset of the validated respondents we can assess the reliability of the validation process, which is very high. For the original 7 point validation scale (which is reduced to a five point scale based on additional variables coding address information matching) the coders gave the same outcome 94.8% of the time for the face-to-face survey and 96.4% of the time for the internet panel.

We can also assess the extent to which the validation suffers from false positive rates. It is plausible to assume that most of our false positives (people who didn't vote, but who we validate as actually having voted) will report not having voted (as only around 25% of non-voters appear to misstate their turnout). Within the group of respondents who report not having voted, 1.5% on the face-to-face are validated as actually having voted and 1.95% on the panel validate as actually having voted. Given that some of these people may have actually mis-recalled in their self-reported turnout, we feel confident that the false positive rate in vote validation is under 2%.

The false negative rate (those who we code as not having voted but actually did vote) is harder to quantify because vote validation is done precisely because we do not fully trust the recall of respondents who claim to have voted. Our coders reported very few problems with determining whether someone on the register had voted (i.e. the marks were generally not ambiguous). We therefore feel confident that there are very few false negatives among those who we confirmed as being registered.

Instead, the major of uncertainty comes from when we have to determine whether someone is not registered or if we are merely unable to locate them. There is therefore some possibility of false negatives in terms of registration and voting within the group we think are not registered. However, for the most part we have erred on coding these cases as "not enough information" rather than not registered where there was any ambiguity in the available information.

		Pre-election		Campaign	
		DNV	**Voted**	**DNV**	**Voted**
Conservative	Freq	84	401	83	388
	%	17.32	82.68	17.62	82.38
Labour	Freq	133	580	127	559
	%	18.65	81.35	18.51	81.49
Liberal Democrat	Freq	30	119	21	83
	%	20.13	79.87	20.19	79.81
SNP	Freq	35	254	37	233
	%	12.11	87.89	13.7	86.3
Plaid Cymru	Freq	1	12	1	15
	%	7.69	92.31	6.25	93.75
UKIP	Freq	52	154	46	167
	%	25.24	74.76	21.6	78.4
Green Party	Freq	22	93	26	113
	%	19.13	80.87	18.71	81.29
Refuse/DK	Freq	34	113	51	149
	%	23.13	76.87	25.5	74.5
Total	Freq	521	1,752	534	1,735
	%	22.92	77.08	23.53	76.47

TABLE B2 VALIDATED TURNOUT RATES AMONG PRE-ELECTION AND CAMPAIGN WAVE RESPONDENTS TO THE BES INTERNET PANEL

		Face-to-face		Panel	
		DNV	**Voted**	**DNV**	**Voted**
Conservative	Freq	57	493	63	443
	%	10.36	89.64	12.45	87.55
Labour	Freq	64	405	105	656
	%	13.65	86.35	13.8	86.2
Liberal Democrat	Freq	9	100	21	142
	%	8.26	91.74	12.88	87.12
SNP	Freq	11	42	37	282
	%	20.75	79.25	11.6	88.4
Plaid Cymru	Freq	0	10	1	12
	%	0	100	7.69	92.31
UKIP	Freq	22	136	32	176
	%	13.92	86.08	15.38	84.62
Green Party	Freq	9	30	16	110
	%	23.08	76.92	12.7	87.3
Refuse/DK	Freq	8	40	1	5
	%	16.67	83.33	16.67	83.33
Total	Freq	182	1,266	278	1,846
	%	12.57	87.43	13.09	86.91

TABLE B3. VALIDATED TURNOUT RATES IN FACE-TO-FACE AND PANEL FOR THOSE REPORTING HAVING VOTED FOR EACH PARTY AFTER THE ELECTIONS

Appendix C - Full comparisons of full sample and voter sample for BES face-to-face and online panel surveys

Turnout

	Face-to-face	Panel	Dif
Non-voter	26.6	8.3	18.3
Voter	73.3	91.2	17.8
Don't know	0.1	0.6	0.5
Total error			36.6

Party member

	Face-to-face	Panel	Dif
Member	3.5	8.2	4.7
Used to be member	4.1	9.2	5.0
Never member	91.9	80.2	11.7
Don't know	0.5	2.4	2.0
Total error			23.4

Campaign interest

	Face-to-face	Panel	Dif
Not at all interested	10.1	3.9	6.2
Not very interested	18.3	7.1	11.2
Somewhat interested	35.6	31.7	3.8
Very interested	36.1	57.3	21.2
Total error			42.4

TABLE C1. COMPARISONS OF POLITICAL ENGAGEMENT VARIABLES BETWEEN THE BES INTERNET PANEL AND FACE-TO-FACE SURVEYS

Party ID

	Total sample			Voters only		
	Face-to-face	Panel	Dif	Face-to-face	Panel	Dif
Conservative	30.1	29.5	0.6	35.5	30.9	4.6
Labour	30.6	30.8	0.2	31.5	31.9	0.4
Lib Dem	6.0	9.1	3.1	6.4	9.3	2.9
SNP	3.4	3.2	0.2	3.7	3.4	0.4
Plaid Cymru	0.5	0.3	0.2	0.5	0.3	0.2
UKIP	4.4	4.8	0.4	4.7	4.9	0.2
Green	2.1	2.6	0.5	2.2	2.6	0.4
BNP	0.0	0.1	0.1	0.0	0.1	0.1
Other	2.6	0.9	1.7	2.9	0.9	2.0
None/No	15.6	14.9	0.7	9.5	12.6	3.1
Don't know	4.7	3.8	0.9	3.1	3.1	0.0
Total error			8.3			14.2
Con lead	-0.5	-1.3	0.7	3.9	-1.0	5.0

Subjective social class

	Face-to-face	Panel	Dif	Face-to-face	Panel	Dif
No class	33.6	21.5	12.1	32.6	19.1	13.6
Middle class	18.0	21.8	3.9	20.9	24.1	3.2
Working class	44.8	50.3	5.5	43.5	51.2	7.7
Other class	1.8	1.8	0.0	1.6	2.1	0.5
Don't know	1.7	4.5	2.8	1.5	3.6	2.1
Total error			24.3			27.1

Private school

	Face-to-face	Panel	Dif	Face-to-face	Panel	Dif
Yes	10.2	15.6	5.5	11.2	15.8	4.6
No	89.8	84.4	5.5	88.8	84.2	4.6
Total error			11.0			9.2

Age

	Face-to-face	Panel	Dif	Face-to-face	Panel	Dif
18-25	13.2	13.9	0.7	10.1	13.0	2.9
26-35	17.2	17.3	0.1	14.0	16.3	2.4
36-45	15.6	15.2	0.4	14.7	15.2	0.5
46-55	18.2	17.7	0.5	19.0	18.1	0.9
56-65	14.8	20.8	6.0	17.0	21.5	4.5
65+	21.0	15.0	6.0	25.2	15.9	9.4
Total error			13.7			20.6

Income

	Face-to-face	Panel	Dif	Face-to-face	Panel	Dif
<40,000	47.1	53.3	6.3	44.7	53.1	8.4
40-000-50,000	8.9	8.6	0.3	9.6	8.8	0.8
50,000-100,000	13.5	11.5	2.0	15.5	12.0	3.5
100,000+	4.0	1.9	2.2	5.0	1.9	3.1
DK/refused	26.6	24.7	1.8	25.2	24.2	1.0
Total error			12.5			16.8

	Working status					
	Face-to-face	Panel	Dif	Face-to-face	Panel	Dif
Full time	45.4	41.5	3.9	45.6	41.6	4.0
Part time	14.0	14.4	0.4	13.9	14.2	0.3
Unemployed	2.7	3.6	0.8	1.8	3.4	1.6
Student	3.8	6.3	2.5	3.0	6.1	3.1
Out of work (other)	9.8	9.8	0.1	7.0	9.6	2.6
Retired	24.0	21.7	2.3	28.5	22.7	5.7
Other	0.3	2.7	2.4	0.3	2.4	2.1
Total error			12.3			19.5

TABLE C2. COMPARISONS OF VARIABLES BETWEEN FACE-TO-FACE AND PANEL SURVEYS